PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The performance of universal early warning scores in different patient subgroups and clinical settings: A systematic review
AUTHORS	Alhmoud, Baneen; Bonnici, Tim; Patel, Riyaz; Melley, Daniel; Williams, Bryan; Banerjee, Amitava

VERSION 1 – REVIEW

REVIEWER	Damian Roland Leicester Hospitals and University I am a researcher interested in PEWS
REVIEW RETURNED	04-Nov-2020

GENERAL COMMENTS	This is a relevent study on EWS from an experience team. The search strategy appears comprehensive (pending a comment below) and the amalagamation of results clear. Two questions: "non-standard EWS developed for a specific subgroup" Could the authors expand on this. How is this different from a specific EWS to be validated in a specific setting i.e what is non-standard? " We were concerned with the use of general EWS in particular patient subgroups and did not assess EWS developed specifically for particular subgroups or settings" why was this, could the authors given an example of a sub-group they are worried about and why that couldn't have been part of this search? This needs clarifying because in the conclusion the authors state "Early warning scores in specific patient subgroups and settings require further prospective validation of their performance in detecting worsening patient outcomes" Do they mean patient subgroups or disease subgroups. I think that conclusion is true but that's not what this study set out to do according to their limitations.

REVIEWER	Nicolás Ariel Grisolía Hospital General de Niños Pedro de Elizalde
REVIEW RETURNED	15-Nov-2020

GENERAL COMMENTS	Comments about the manuscript: "The performance of early warning scores in different patient subgroups and clinical settings: A systematic review" The topic of early warning scales is very interesting, not because of the novelty, but because of the current approach in line with advances in quality and patient safety. I Suggest the following modifications: 1- I Suggest incluide pediatric patients like a new subgroup of patients, because increase the external validity 2- About Section "data sources" i suggest limiting the study period to 10 years. Scientific and technological changes can generate a bias in old methods with respect to current approaches, with erroneous conclusions 3- You must explain in "discussion" section, the reason of the exclussion of many papers. 4- About "Outcome" you must define what you mean by "mortality", "cardiac arrest", "ICU transfer" to increase the external validity of the study 5- Tables must be self-explanatory, abbreviations must be clarified in all tables (Figure 3, Figure 4). Table 1 out of range, Table 2 out of range Page 30, Table 4: Abbreviations are missing. Page 38: I suggest explaining x-axis and y-axis. The colors do not match, the table is not well understood. Page 40: table out of margin. I suggest explaining "x axis" and "y axis" 6- I suggest putting a baseline at value 0.5 in AUC, for a better comparission 7- Page 21 - 26. Table 2. Adjust table, modify margins. Modify character axis in "settings", "Study design", "EWS" "author, year" 8- Page 17 Figure 1. The numbers do not match. Redo the flow chart 9- Page 18 Figure 2. The figure is not easy to understand. Redo it 10- Page 19. Figure 3. The leggend doesn´t match on the same page 11- Figure 2 and figure 3. (Page 19 and 20). It is not appropriate to put a histogram in this case, it is better to leave the dot plot because it is a descriptive study. I suggest doing a meta-analysis to extract a summary measure and then make a similar graph (forest plot) 12- Figure 1. Page 17. "Search strategy diagram using PRISMA (Preferred reporting items for systematic reviews and meta-analyzes)". Remove the content "(Preferred reporting items for systematic reviews and meta-analyzes)" because you are NOT doing a meta-analysis

	13- "Figures and Tables" section. Remove the index. Put the figure of table with your following title and legend 14- "Discussion" (major modification) - It's not extensive - Does not explain about the AUC analysis measure - I suggest doing a meta-analysis before drawing conclusions about the heterogeneity of the sample. - It does not speak of the strengths of the study. It does not talk about the strengths of the EWS. It is seen in most of the studies that the AUC is> 0.6, with the majority> 0.7, which is a good discrimination, taking into account the large interval of years taken. - Compared all subgroups equally, I suggest meta-analysis with subgroup analysis to increase internal validity and better conclusions - Compare different scales contributes to the apparent bias. I Suggest standardize scales, or compare in subsets of scales with similar features

REVIEWER	Jacqueline Birks University of Oxford, UK I have previously worked with Tim Bonnici
REVIEW RETURNED	16-Nov-2020

GENERAL COMMENTS	The introduction fails to define an EWS, and there is very little information about EWSs. EWS are a measure of risk of a certain event. EWS are derived from a prediction model, to predict risk of a particular outcome in a patient group based on patient characteristics at baseline, at a time from baseline. An EWS score could be derived from a statistical model in which case the patient's characteristics are entered into the statistical equation to calculation the absolute risk. EWS can also be derived using expert opinion, when a panel of clinical experts decide which patient characteristics are relevant to the EWS, and how the risk is estimated from the different values of the characteristics, and what the weight of each characteristic is. This is essential information but none of this background is in the introduction. There is mention of NEWS , but not even a short description of the aim of NEWS, the patients characteristics that are required, that this is an EWS based on expert opinion and produces an integer score between which is related to risk of deterioration . EWSs need to be validated, to test how well they perform in the relevant population. None of this is described. Validation and methods of validation are not described. Difference between internal and external validation not explained. Then the EWSs needed to be tested in the relevant population. Therefore the authors set out the objective " we aim to describe the performance of EWS in different diseases and different clinical settings." This suggests that the SR will cover every EWS in every situation. There needs to be justification for this objective. The authors end up with 105 studies that cover 4 different settings in hospital, several patient populations defined by disease, several outcomes, assessed at many different time points from baseline. My first question is who would find this review useful, who will be looking for universal evidence for any EWS ? The table of all included studies is reported, including the list of

	EWSs that each study evaluates. Some studies evaluate and compare more than one EWS Some of these EWS are specific for outcome or patient group. It would be helpful to have a table describing relevant features of each EWS that is evaluated in at least one of the included studies. It is difficult interpreting the figures. Figure 2 . Number of studies on EWS performance in different subgroups and settings. Are the subgroups, sub groups of patients, and can each subgroup appear in more than one setting? What is this figure telling us? I do not find figures 3 and 4 informative, I cannot interpret the results. Studies of varying methodological quality and accuracy are pooled Rather than attempt a superficial SR of everything it may be better to target a disease group, or setting and deal with the validation in depth. The included studies are very heterogeneous, and of very variable quality. A detailed discussion of a subgroup relating to a single disease would emphasise the differences between studies, allow a discussion of the methodology, and provide a more informative conclusion and discussion.

VERSION 1 – AUTHOR RESPONSE

Reviewer **1**

Comments to the Author

This is a relevent study on EWS from an experience team. The search strategy appears comprehensive (pending a comment below) and the amalagamation of results clear.

Thanks for this kind remark.

1- "non-standard EWS developed for a specific subgroup"

Could the authors expand on this. How is this different from a specific EWS to be validated in a specific setting i.e what is non-standard?

We thank the reviewer for the comment. Non-standard EWS are EWS designed for a specific patient group (rather than the general population) and are usually validated in that group. We have now clarified this issue in introduction and exclusion criteria of the revised manuscript.

2- " We were concerned with the use of general EWS in particular patient subgroups and did not assess EWS developed specifically for particular subgroups or settings"

Why was this, could the authors given an example of a sub-group they are worried about and why that couldn't have been part of this search?

This needs clarifying because in the conclusion the authors state "Early warning scores in specific patient subgroups and settings require further prospective validation of their performance in detecting worsening patient outcomes"

Thanks for this helpful remark. Pragmatically, EWS developed for use in general populations have been used and are used in particular patient subgroups.

In the discussion of the revised manuscript, we mention EWS with standardised parameters and thresholds which are implemented in all disease subgroups where they may not be

applicable. CVD represents a patient subgroup of concern with only one study found, which we have explained in the discussion section of the revised manuscript.

In the conclusion, we have changed our text as follows:

"Early warning scores developed for general patient populations require further validation of their performance for detecting worsening outcomes in specific disease subgroups and settings."

3- Do they mean patient subgroups or disease subgroups. I think that conclusion is true but that's not what this study set out to do according to their limitations.

We feel that inevitably "patients" and "diseases" overlap and it is hard to distinguish the two. We meant both in terms of subgroups, and have now clarified in the revised manuscript. The more important distinction we try to make is the "disease/patient" subgroup compared with the "clinical setting" subgroup.

Reviewer 2

The topic of early warning scales is very interesting, not because of the novelty, but because of the current approach in line with advances in quality and patient safety.

I Suggest the following modifications:

1- I Suggest include pediatric patients like a new subgroup of patients, because increase the external validity

Thanks for this remark. EWS designed for adult patients are distinct from paediatric EWS and were beyond the scope of our review from the outset. The study aims to evaluate the applicability of standardised EWS in all adult settings where they are to be implemented. Therefore, we have excluded paediatric patients and paediatric EWS as this was not in the scope of our study.

2. About Section "data sources" i suggest limiting the study period to 10 years. Scientific and technological changes can generate a bias in old methods with respect to current approaches, with erroneous conclusions

Thanks to the reviewer for this comment. Our search results yielded studies included from 2006 (14 years period). We included few studies from 2006, 2007 and 2009 in specific disease subgroups since limited researches was done during this period. We feel it is useful to include this in our study to show increased interest and research in EWS over time.

3. You must explain in "discussion" section, the reason of the exclussion of many papers.

We have explained this to some extent in the figure and also in the discussion of the revised manuscript.

4. About "Outcome" you must define what you mean by "mortality", "cardiac arrest", "ICU transfer" to increase the external validity of the study

Thanks-this is now addressed in the manuscript.

5. Tables must be self-explanatory, abbreviations must be clarified in all tables (Figure 3, Figure 4).

We have now addressed these issues in the revised manuscript.
Table 1 out of range, Table 2 out of range

This was a submission error- apologies for this oversight which we have now corrected.

Page 30, Table 4: Abbreviations are missing.
We have now added abbreviations in the revision.

Page 38: I suggest explaining x-axis and y-axis. The colors do not match, the table is not well understood.

Thank you-we have now explained and fixed these issues.

Page 40: table out of margin. I suggest explaining "x axis" and "y axis"
Thank you-we have now explained and fixed these issues.

6. I suggest putting a baseline at value 0.5 in AUC, for a better comparission

We have now addressed this issue.

7. Page 21 - 26. Table 2. Adjust table, modify margins. Modify character axis in "settings", "Study design", "EWS" "author, year"

Thanks-we have made these suggested changes.

8- Page 17 Figure 1. The numbers do not match. Redo the flow chart

Thank you-we have now fixed this issue.

9- Page 18 Figure 2. The figure is not easy to understand. Redo it

Thank you-now done.

10- Page 19. Figure 3. The legend doesn´t match on the same page

Thank you-now addressed in the revised manuscript.

11- Figure 2 and figure 3. (Page 19 and 20). It is not appropriate to put a histogram in this case, it is better to leave the dot plot because it is a descriptive study. I suggest doing a meta-analysis to extract a summary measure and then make a similar graph (forest plot)

Thanks for the suggestions, which we have adopted in the revision. A meta-analysis was done for mortality in all subgroups.

12- Figure 1. Page 17. "Search strategy diagram using PRISMA (Preferred reporting items for systematic reviews and meta-analyzes)". Remove the content "(Preferred reporting items for systematic reviews and meta-analyzes)" because you are NOT doing a meta-analysis

Now addressed in the revision.

13- "Figures and Tables" section. Remove the index. Put the figure of table with your following title and legend

Now addressed in the revision.

14- "Discussion" (major modification)

- It's not extensive

We have extended the discussion section.
- Does not explain about the AUC analysis measure

We have added this in the revision.
- I suggest doing a meta-analysis before drawing conclusions about the heterogeneity of the sample.

We have conducted a meta-analysis for all EWS in all subgroups.

- It does not speak of the strengths of the study. It does not talk about the strengths of the EWS. It is seen in most of the studies that the AUC is> 0.6, with the majority> 0.7, which is a good discrimination, taking into account the large interval of years taken.

Thanks for these excellent recommendations. We have added these improvements to the discussion and results sections.

- Compared all subgroups equally, I suggest meta-analysis with subgroup analysis to increase internal validity and better conclusions

Thanks-now done in the revision.

- Compare different scales contributes to the apparent bias. I Suggest standardize scales, or compare in subsets of scales with similar features

Thanks for these suggestions. Results of studies are compared when AUC was the measure used. Other studies were included to show heterogeneity in measures used.

Reviewer 3

The introduction fails to define an EWS, and there is very little information about EWSs. EWS are a measure of risk of a certain event. EWS are derived from a prediction model, to predict risk of a particular outcome in a patient group based on patient characteristics at baseline, at a time from baseline. An EWS score could be derived from a statistical model in which case the patient's characteristics are entered into the statistical equation to calculation the absolute risk. EWS can also be derived using expert opinion, when a panel of clinical experts decide which patient characteristics are relevant to the EWS, and how the risk is estimated from the different values of the characteristics, and what the weight of each characteristic is. This is essential information but none of this background is in the introduction.

Thanks. We have used definitions from literature and improved the text as suggested.

There is mention of NEWS , but not even a short description of the aim of NEWS, the patients characteristics that are required, that this is an EWS based on expert opinion and produces an integer score between which is related to risk of deterioration .

This is now added to the new revised manuscript.

EWSs need to be validated, to test how well they perform in the relevant population. None of this is described. Validation and methods of validation are not described. Difference between internal and external validation not explained. Then the EWSs needed to be tested in the relevant population.

We have addressed all these issues further in the revised manuscript.

Therefore the authors set out the objective:
" we aim to describe the performance of EWS in different diseases and different clinical settings."
This suggests that the SR will cover every EWS in every situation. There needs to be justification for this objective.

We have clarified these parts in the revision.

The authors end up with 105 studies that cover 4 different settings in hospital, several patient populations defined by disease, several outcomes, assessed at many different time points from baseline.
My first question is who would find this review useful, who will be looking for universal evidence for any EWS ?

We have clarified this in the revised manuscript. The benefit of our systematic review is to provide an overall insight into the performance of EWS in specialised subgroups to foresee the outcome if implemented in such settings and to know if current studies are sufficient to recommend this implementation. The alternative would be to look at each setting in individual reviews which seems wasteful in terms of effort.

The table of all included studies is reported, including the list of EWSs that each study evaluates. Some studies evaluate and compare more than one EWS
Some of these EWS are specific for outcome or patient group. It would be helpful to have a table describing relevant features of each EWS that is evaluated in at least one of the included studies.

A table of different EWS features is included in the appendix. All included EWS are not specific to a disease group; they are standardised EWS used for all patient groups. The EWS predict critical outcomes as cardiac arrest, ICU transfer and mortality but not specific disease-related outcomes.

It is difficult interpreting the figures.
Figure 2 . Number of studies on EWS performance in different subgroups and settings.
Are the subgroups, sub groups of patients, and can each subgroup appear in more than one setting?
What is this figure telling us?

We have re-done figures in the revision with improved explanations.

I do not find figures 3 and 4 informative, I cannot interpret the results. Studies of varying methodological quality and accuracy are pooled
Rather than attempt a superficial SR of everything it may be better to target a disease group, or setting and deal with the validation in depth.

Thanks for these remarks. Our systematic review and meta-analysis looked at all subgroups using different EWS and with NEWS. We set out to give an overview of the use of EWS. Figure 3 and 4 have been redone. The CVD subgroup is discussed in the discussion section of the revision.

The included studies are very heterogeneous, and of very variable quality. A detailed discussion of a subgroup relating to a single disease would emphasise the differences between studies, allow a discussion of the methodology, and provide a more informative conclusion and discussion

Many thanks for the comments. We politely disagree. Given the widespread recommendation and use of general or standard EWS, we felt that a knowledge gap was an overview of the use of standard, general EWS in different subgroups and the overall research in this area. The study aims to show the level of heterogeneity between studies in subgroups. We have tried to clarify this in the revision.

VERSION 2 – REVIEW

REVIEWER	Nicolás Ariel Grisolía Hospital General de Niños "Pedro de Elizalde". Argentina, Ciudad Autónoma de Buenos Aires
REVIEW RETURNED	11-Jan-2021

GENERAL COMMENTS	Hi! Congratulations. Thanks for accept my suggestions for you. One last suggest: Page 25 and page 26 Forest plots. They can´t be read in detail. Adjust the scales of the both plots.

REVIEWER	Jacqueline Birks University of Oxford, UK I have previously worked with Dr Bonnici
REVIEW RETURNED	14-Jan-2021

GENERAL COMMENTS	There is no doubt that this was a difficult and challenging systematic review to carry out. There are numerous studies, some of which are not of the highest standard, each of which requires careful reading. There is still a lack of clarity in the text which partly arises because of confusing nomenclature. EWS is an acronym for early warning scores, any early warning score, nothing specific. Confusion arises when EWS is then used, as here, to cover all the EWS studied in this systematic review which are a specific subset of EWS, a subset developed with the objective of being useful for all patients in all disease groups in all clinical settings. This needs to be clear. It would be helpful to have a different acronym for this subgroup. Sometimes the authors refer to this subset of EWS as standardised scores, but they are not standardised scores. Is standardised the correct description? In what way is the NEWS standardised? I do not think they are standardised unless the authors are referring to the use of a standard set of predictors. Would general be a better term, or universal? Whatever the term, there should be a definition, with acronym, and the authors should justify this definition. In addition the inclusion criteria for this subset of EWS should use this definition. Title The performance of early warning scores in different patient subgroups and clinical settings: A systematic review It is not clear from the title that this SR is only looking at a specific subset of EWS. Introduction The authors state “Across diseases, patient deterioration can range from critical care

	review and sepsis, to cardiorespiratory arrest and death, resulting in strain on healthcare resources(1,2)." Why is the first cause for concern the use of health care resources, and why is it a strain? It is axiomatic that all patient illness uses resources. Data analysis – The whole section needs editing as it does not make sense. Quality assessment "Almost half of the studies (n=49; 48%) validated in <500 patients with either multiple observations or a single observation set (Tables 1 and 2)." The authors do not explain what multiple observations mean. They are making reference to the fact that patients may have more than one set of observations in a data set, but nowhere do they discuss this possibility and the effect it may have on a validation study. Multiple observations are mentioned again in the discussion. Results. Methodology The authors state: "Longer-term outcomes following EWS implementation were assessed in five studies in ICU, medical and surgical settings. Results were mixed. Mortality rate was reduced in three of the studies: in ICU(8) and medical settings(66); and no improvement was observed in a medical setting. However, the study duration was likely to be inadequate, e.g. four months(67). The ICU transfer and cardiac arrest rates improved in surgical(68) and medical settings(66), but deteriorated in another study in a medical setting(67). " Here the authors are describing the testing of EWS in clinical settings. These studies are not validation studies. Testing comes after validation. These studies are not included studies and it is confusing to start discussing these studies. These studies appear at the bottom of Table 2.S and a footnote should explain why they are there. Discussion The authors state: "Inconsistency in evaluation and the lack of high-quality validation ultimately affects how EWS can and should be used in clinical practice, e.g. predicting risk of future deterioration versus actual deterioration(29)." Not clear what this means as EWS are meant to predict risk of future deterioration. Table S2. Hamtology must be a mis spelling,

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3
Ms. Jacqueline Birks, University of Oxford
Comments to the Author:
There is no doubt that this was a difficult and challenging systematic review to carry out. There are numerous studies, some of which are not of the highest standard, each of which requires careful reading.
There is still a lack of clarity in the text which partly arises because of confusing nomenclature.

EWS is an acronym for early warning scores, any early warning score, nothing specific. Confusion

arises when EWS is then used, as here, to cover all the EWS studied in this systematic review which are a specific subset of EWS, a subset developed with the objective of being useful for all patients in all disease groups in all clinical settings. This needs to be clear. It would be helpful to have a different acronym for this subgroup. Sometimes the authors refer to this subset of EWS as standardised scores, but they are not standardised scores. Is standardised the correct description? In what way is the NEWS standardised? I do not think they are standardised unless the authors are referring to the use of a standard set of predictors. Would general be a better term, or universal? Whatever the term, there should be a definition, with acronym, and the authors should justify this definition.
In addition the inclusion criteria for this subset of EWS should use this definition.

- **Many thanks for this very helpful comment. We agree that standardised is not the right term. We have opted, as suggested, for "universal" EWS with general use. We did not define a further acronym for universal EWS to avoid confusion.**
- **The type of EWS chosen in the systematic review falls under the category of "standardized" prediction tools when developed and validated. The term is used in a number of articles on EWS referring to the unified use of a single EWS with a set of parameters to predict deterioration in any patient subgroup, e.g NEWS, NEWS 2, and SEWS(1)(2). By "standardised", we mean "to make things of the same type all have the same basic features" or "to compare one thing to something accepted as a model". We defined "universal" in this context as: "applicable everywhere or in all cases; or that can be in many places at the same time". Therefore, universal EWS are those "having a consistently measured way of assessing a patient"(3).**
  The term has been defined with other technical terms used in the paper.

  The criteria used for inclusion is also clarified using the appropriate terms.

Title
The performance of early warning scores in different patient subgroups and clinical settings: A systematic review
It is not clear from the title that this SR is only looking at a specific subset of EWS.

- **The title is now modified to fit with the type of EWS used:**
  "The performance of universal early warning scores in different patient subgroups and clinical settings: A systematic review."

Introduction
The authors state
"Across diseases, patient deterioration can range from critical care review and sepsis, to cardiorespiratory arrest and death, resulting in strain on healthcare resources (1,2)."
Why is the first cause for concern the use of health care resources, and why is it a strain? It is axiomatic that all patient illness uses resources.

- **Thanks for this remark. We have removed the phrase "resulting in strain on healthcare resources". We have added a sentence to explain our logic:**
  "For example, the 20,000 in-hospital cardiac arrests per year in England are associated with costs of £50 million for resuscitation and post-arrest care(4). Around the world, earlier recognition and prevention of deterioration in unwell patients has far-reaching implications for reduction in mortality and morbidity, reduction in the cost of

healthcare, and allocation of scarce high dependency and critical care resources. Preventive interventions are needed to overcome these challenges (5).”

Data analysis –
The whole section needs editing as it does not make sense.

- **Thank you for this notice, the section is now written clearly.**

Quality assessment
“Almost half of the studies (n=49; 48%) validated in <500 patients with either multiple observations or a single observation set (Tables 1 and 2).”
The authors do not explain what multiple observations mean. They are making reference to the fact that patients may have more than one set of observations in a data set, but nowhere do they discuss this possibility and the effect it may have on a validation study. Multiple observations are mentioned again in the discussion.

- **The effect of using single and multiple observation is now clarified in the discussion.**

The authors state:
“Longer-term outcomes following EWS implementation were assessed in five studies in ICU, medical and surgical settings. Results were mixed. Mortality rate was reduced in three of the studies: in ICU(8) and medical settings(66); and no improvement was observed in a medical setting. However, the study duration was likely to be inadequate, e.g. four months(67). The ICU transfer and cardiac arrest rates improved in surgical(68) and medical settings(66), but deteriorated in another study in a medical setting(67). “
Here the authors are describing the testing of EWS in clinical settings. These studies are not validation studies. Testing comes after validation. These studies are not included studies and it is confusing to start discussing these studies.

- **Thank you, we agree to this point. It will cause confusion since it was not part of the detailed evaluation of validation studies. Therefore, these studies were removed and the section in results is also omitted.**

These studies appear at the bottom of Table 2.S and a footnote should explain why they are there.

- **These are removed from the table now.**

“Inconsistency in evaluation and the lack of high-quality validation ultimately affects how EWS can and should be used in clinical practice, e.g. predicting risk of future deterioration versus actual deterioration (29).”
Not clear what this means as EWS are meant to predict risk of future deterioration.

- **Thank you, it is meant to show the role of EWS in early prediction for short term outcomes rather than long term outcomes. We removed this line to keep the focus on the EWS role.**

Table S2. Hamtology must be a mis spelling,

- **Thank you. The spelling is corrected now.**

VERSION 3 – REVIEW

REVIEWER	JACQUELINE BIRKS University of Oxford, UK I have previously worked with Dr Bonnici
REVIEW RETURNED	04-Mar-2021

GENERAL COMMENTS	This is an important review which summarises the results of many studies, and provides a critical discussion of the universal EWS. The authors have addressed my comments and answered my questions.